Power-law distribution of individual Hirsch indices, the comparison of merits in different fields, and the relation to a Pareto distribution

Pekka Pyykkö[1]

[1]*Department of Chemistry, University of Helsinki, P.O. Box 55 (A.I. Virtasen aukio 1), FI-00014 Helsinki, Finland*

E-mail: Pekka.Pyykko@helsinki.fi

A data set of Hirsch indices, h, for Finnish scientists in certain fields is statistically analyzed and fitted to $h(n) = Pn^p$ for the n-th most-quoted scientist. The precoefficient P is characteristic for the field and the exponent p is about -0.2 for all data sets considered. For Physics, Chemistry and Chemical Engineering, the P are 49.7(8), 41.3(6), and 21.4(6), respectively. These p values correspond to Pareto exponents of about -7 for the distribution of Hirsch indices h.

PACS numbers: 01.30.-y

I. INTRODUCTION

The Hirsch index h [1, 2] provides a rough but robust measure on the total citation impact of an individual, until the time of observation. More exactly it means having h papers, each cited at least h times. In addition to persons, it also can be defined for universities, journals etc. The values are very different for different fields and the question is, how to compare the values between fields?

We had available a small data set of the h values in Chemistry, Physics, and Chemical Engineering for Finnish scientists. A statistical study reveals an interesting power-law distribution and gives a hint on the relative weighting factors that may apply between different fields.

II. METHOD AND RESULTS

The data were determined from the ISI Web of Knowledge using the data set in General Search from 1945 onwards. This database only contains references in journals to papers in journals. Most data points were obtained in November 2005. The most-quoted one-third of the points inside each area, k, was fitted using Gnuplot to

$$h(n) = Pn^p \tag{1}$$

where $h(n)$ is the h of the n:th-most quoted scientist, P is a precoefficient and p is an exponent, found to be surprisingly constant between different fields. The obtained values are shown in Table I and the quality of the fits is demonstrated in Fig. 1. The figures in parentheses give the asymptotic standard error. In this data set, for the given country at the given time, the workers in different areas mostly share the same background and general working conditions, like the typical research-group size and budget. Assuming that they also are equally gifted and hard-working, we then suggest that the ratios of P between different fields would form a possible basis for comparing scientific merit between fields.

Podlubny[3] recently compared the total numbers of citations in various fields in United States. He found them to be fairly constant from 1992 to 2001 and suggested that they would form a useful normalization factor for comparing individual scientific performance between fields.

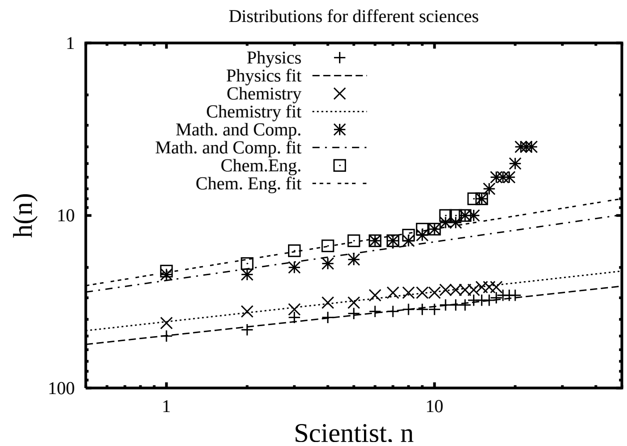

FIG. 1: The fits (1) for Physics, Chemistry, Mathematics plus Computer Science, and Chemical Engineering. For the two latter fields the entire data sets are shown as points, although the fits only include the k highest points in Table I.

TABLE I: The fits for certain areas. k is the number of points included in the fit. All data refer to Finland.

Area	k	P	p
Medicine	4	90(3)	-0.22(3)
Bio/eco	5	59(4)	-0.23(7)
Physics	14	49.7(8)	-0.169(9)
Chemistry	17	41.3(6)	-0.173(7)
Math and Comp	8	23.8(1.5)	-0.22(5)
Chem. Eng.	5	21.4(6)	-0.25(3)

In Table II we compare the present relative $P_{\rm rel}$ factors (with Physics normalized to 1) to the square roots of Podlubny's relative citation numbers. An average of his 1992-2001 data is used.

Recall here that the lower limit for the total number

TABLE II: The relative prefactors, $P_{\rm rel}$, with Physics normalized to one and the square roots of the number of total citations, $(C_{\rm rel})^{1/2}$, with Physics normalized to one.

Area	$P_{\rm rel}$	$(C_{\rm rel})^{1/2}$
Medicine	1.8	2.0
Bio/Eco	1.2	
Physics	1	1
Chemistry	0.83	0.88
Math and Comp	0.48	0.23
Chem. Eng.	0.43	

of citations, $N_{c,tot} = h^2$ and a typical number is[2]

$$N_{c,tot} = ah^2, \tag{2}$$

with a about 3-5 [2].

III. FURTHER DATA SETS

A list of the h values for 40 'Dutch' chemists was published by Faas[4]. Both people of Dutch origin anywhere in the World, and people from anywhere, working in The Netherlands were included. As seen from Fig.2, all points fit well the values P = 105.5(2.4), p = -0.212(11).

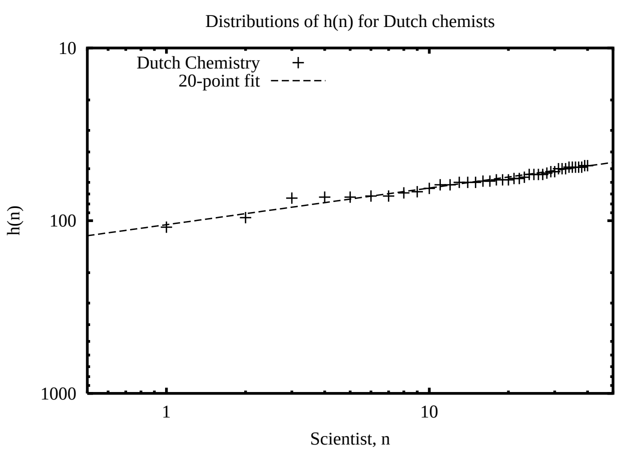

FIG. 2: The h values of forty 'Dutch' chemists from Faas[4]. The line is fitted to the points 1-20 and has p = -0.212(11).

IV. RELATION TO THE PARETO DISTRIBUTION

In economic theory, V. Pareto found in 1896 [5] the number of holders of income I in a country to scale for high incomes as I^x, with x about -2 ([5], see ref. [6, 7]). The same law was found by Zipf to hold for word frequencies in linguistics and by Lotka for numbers of papers among authors[8]. It is known in many other fields, like size distributions of cities in a country, earthquakes, wars etc. [7].

From eq. (1),

$$n(h) = (h/P)^{1/p}. \tag{3}$$

Introducing the density of individuals per unit of h, $N(h)$,

$$n = \int_{h_n}^{\infty} N(h')dh', \tag{4}$$

we can interprete $N(h)$ as the derivative

$$N = -dn/dh. \tag{5}$$

Then, using eq.3,

$$N(h) = P^{-1/p} h^{\frac{1}{p}-1}. \tag{6}$$

For the Finnish p for Physics and Chemistry, the corresponding Pareto exponent x would become -6.9 and -6.8, respectively.

The main conclusions are that the P value for Chemical Engineering is about half of that for Chemistry, and that the p values for the data sets considered are about -0.2.

Acknowledgments

Claus Montonen pointed out ref. [3].

[1] J. E. Hirsch, arXiv:physics/0508025 v5, 29 Sep 2005.
[2] J. E. Hirsch, Proc. Natl. Acad. Sci. (USA) **102**, 16569 (2005).
[3] I. Podlubny, J. Scientometrics **64**, 95 (2005). arXiv math ST/0410574.
[4] F. Faas, Chemisch 2 Weekblad (March 11, 2006), p. 18.
[5] V. Pareto, *Cours d'Economie Politique*, Droz, Genève (1896). As quoted in ref. [7].
[6] J. D. Farmer, M. Shubik, and E. Smith, Physics Today **58**, No. 9, 37 (2005).
[7] M. E. J. Newman, Contemporary Physics **46**, 323 (2005).
[8] A. J. Lotka, J. Washington Acad. Sci., **16**, 317 (1926).